

September 11, 1963

Mr. Edward Rosse
Room 322, Administration Wing
OASI - Data Processing Division
Baltimore 35, Maryland

Dear Mr. Rosse:

Over the past several months we have been trying to establish working relationships with private and government agencies responsible for large, machinable files of statistical data characterizing the U. S. population. You probably know of Dr. H. B. Newcombe's work on record linkage (on Canadian data), motivated by the importance of family correlations for genetic analysis. There seems little doubt now that OASI records will be the focal point of record linkage in the future; I hope the opportunity will be developed to do this for the benefit of demographic and health-oriented statistics as well as for the other governmental functions of tax and benefits administration.

Gene Garfield suggested that I contact you for information on the long range strategy of such programs, to the extent they have been formulated. For example, I note that OASI numbers are now recorded on (California) death certificates. Should I infer that OASI receives copies of the death records for matching its own files?

My own studies have to do with fertility patterns ('childspacing') and related problems, including a somewhat surprising social stratification of seasonal timing of births -- and I would also predict of marriages and deaths. I feel these seasonal cycles may on the one hand distort, on the other, enrich demographic studies (if they can be objectively analyzed), in addition to their inherent interest. Furthermore, detailed dating of records should be invaluable for record checking and linkage -- e.g. if birth dates (when available) appeared on death certificates.

At present I have one tangible problem which you might help to answer. We have some indications of 'discrepant' birth month distributions among (1) the propositi of Terman's study of genius and (2) different blood groups of donors at a San Francisco blood bank. But, incredibly, there is not convenient (detailed) source of control birth dates of extant individuals, except probably the OASI files. These should have additional structure of great interest -- occupational and income classification. Perhaps there are even measures of family linkage! Both for the immediate problem and for more extended interests, may I impose on you for information on (1) the organization of your machined records, and (2) the conditions and feasibility of enlisting OASI cooperation for such studies, on samples in the range of 10^5 and 10^6 .

W. S. Garfield
OASI
Rec'd

I have been in touch with Vital Statistics (natality) and Census (population) on these questions (and have an application to NIH in process for funding these studies). But Census unfortunately has not usually carried its files to the requisite detail and Vital Statistics can only refer to new events.

We are well set up for extensive computation here, with an IBM-7090 and a Burroughs B5000 system. The former is being organized for extensive time-sharing and remote I/O with dynamic, graphic displays, as well as the usual printouts. Some satellite computers (now a PDP-1 at the Computation Center and a small LINC in our own laboratory) are involved in this as terminal and display stations. Programming goes reasonably smoothly. So we are very keen to tie on to the data to test our mechanistic speculations and exploit these facilities.

Sincerely yours,

Joshua Lederberg
Professor of Genetics